

Study of correlations between logfile-based metrics of interaction and the quality of synchronous collaboration

Georgios Kahrmanis

Human-Computer Interaction Group, University of Patras, Greece

kahrmanis@ece.upatras.gr

Irene-Angelica Chounta

Human-Computer Interaction Group, University of Patras, Greece

houren@ece.upatras.gr

Nikolaos Avouris

Human-Computer Interaction Group, University of Patras, Greece

avouris@upatras.gr

Abstract. This paper presents a study that investigates correlations between various metrics of interaction based on interaction events automatically recorded by an online collaboration tool, and collaboration quality as it is assessed by human agents according to a rating scheme. The study concerns a large dataset of synchronous problem-solving technology-enhanced collaborative learning activities.

1 Introduction

Research on technology-enhanced collaborative learning has focused on analyzing and assessing computer-supported collaborative activities with the aid of various analysis and evaluation tools. Collaboration support systems allow automatic loggings of users' actions that are maintained in suitably structured logfiles, which

can then provide the basis for the calculation of automated metrics of interaction. In technology-enhanced collaborative learning, the state of evolving knowledge must be continuously displayed by collaborating participants with each other (Stahl, 2002). Therefore, what one participant communicates with others is accessible to researchers through logfile entries, providing thus an objective source for analysis (Dillenbourg et al., 1995). Such data can then be subjected to automated statistical elaborations which, in the form of suitably implemented metrics, may be useful for indicating important aspects of collaboration quality.

There is, however, limited evidence that automatic metrics of interaction reported in the literature are capable of indicating collaboration quality (e.g. Avouris et al., 2004). There either is no rigorous examination of their indicatory value, or they are compared to indirect indications of collaboration quality such as the quality of the outcome of a collaborative process (Aditomo, & Reimann, 2007).

This study presents an extensive set of metrics of interaction that were designed and implemented in order to be statistically tested for their suitability to indicate core aspects of collaboration quality. Each metric was kept rather simple, so that potential results would be easily interpretable and additional, more sophisticated metrics would be developed in the future, informed by current findings. Correlational statistical tests were then carried out comparing scores that the metrics took in a large dataset with collaboration quality ratings applied following a different methodological approach (Kahrimanis et al., 2009).

2 Automatic metrics of interaction

Metrics designed and developed in the frame of this study were informed by an existent metric set implemented by the Synergo tool, which provides a chat and a shared workspace supporting collaborative modelling activities (Avouris et al., 2004). Like most collaboration support tools, Synergo keeps logs of events of users' interaction with the tool in a logging format inspired by the Object-oriented Collaboration Analysis Framework (OCAF) (Avouris et al., 2003). According to OCAF, collaborative activity can be described in a four-dimensional space, the four axes of which are *time*, *actor*, *object*, and *typology*. *Time* refers to the temporal moment of the occurrence of a users' action, *actor* is the collaboration participant who generated an action, *object* refers to an object created throughout the process (e.g an item in the shared workspace or a chat message), and *typology* contains a characterisation of an event according to some predefined categorisation. In that manner, all Synergo logfile records follow a format based on these core dimensions, and metrics of interaction are calculated taking advantage of the structure of data gathered.

For the purposes of this study, the existent set of metrics of interaction of Synergo was reshaped and significantly augmented. Four categories of events

were defined based on the kind of object that an event relates to: *Chat messages (C)*, *Main actions in the Workspace (MW)* (including only these actions in the workspace that lead to significant changes in the developed model), *Overall actions in the Workspace (OW)* (including actions in the workspace of secondary importance as well, such as the movement or resizing of existent objects), and *overall EVents (EV)* (including all categories of events captured). Generic types of metrics were then defined that involve calculations of the data logged, taking advantage of other information of log annotations, such as the typology of actions, temporal aspects, and interchanges of the actors of events. Eight such types of metrics were then developed, each one of them applied for each category of events mentioned above: *number of []*, *rate of []*, *symmetry of []*, *alternations in []*, *rate of alternations in []*, *mean response time in []*, *median response time in []*, and *number of [] gaps per X (parametric) seconds* (with the square brackets standing for any category of events). 4 additional metrics that could not be covered by the above typology were also added so that the final set used consisted of 36 metrics. The whole metric set is illustrated in Figure 1.

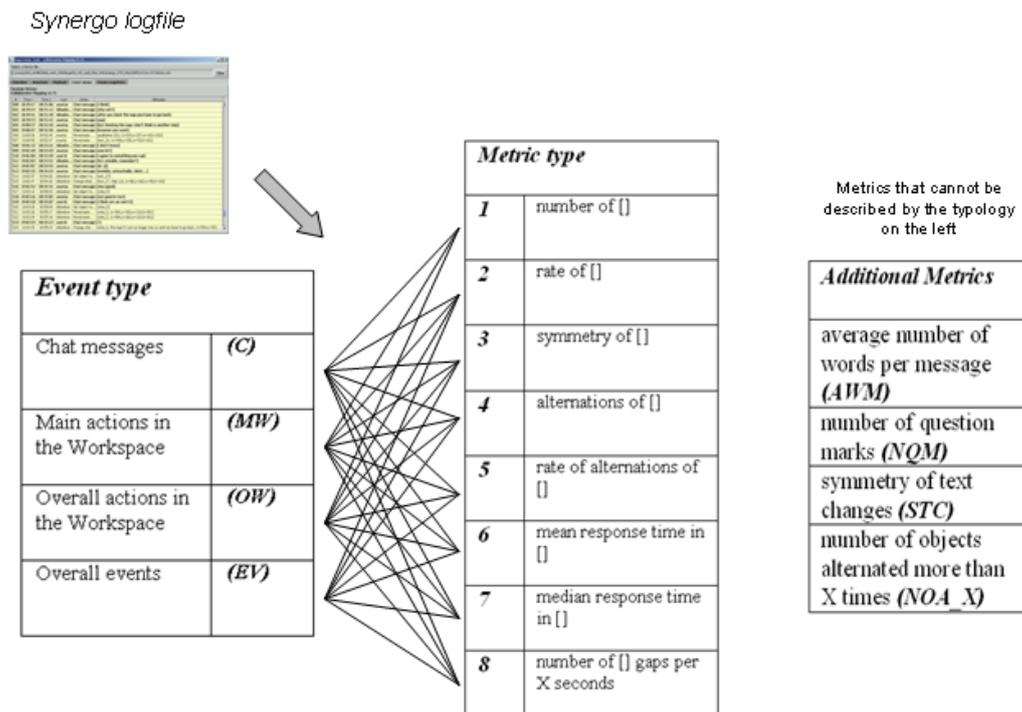


Figure 1. The augmented set of Synergo's automated metrics of interaction

3 Collaborative setting

The objectives of this study implicated that the newly developed metrics should be tested empirically in a large-scale, real-world scenario. Therefore, extended collaborative activities were designed and put on, in order to provide a rich data source for statistical analysis of the values that metrics take in common,

naturalistic CSCL activities. The collaborative activities studied involved approximately 350 university students in the Computer Engineering in the Electrical and Computer Engineering Department of the University of Patras, Greece, engaged in jointly building the diagrammatic representation of an algorithm as an assignment of a two-hour laboratory session that was part of the first-year of studies course “Introduction of Computers and Algorithms”. Students interacted synchronously through Synergo (Avouris et al., 2004), communicating via an integrated chat tool, and jointly designing a flow-chart representation of an algorithm in Synergo’s shared whiteboard. Collaborative sessions lasted from 45 to 75 minutes and students worked in dyads. In order to motivate students to work on the exercises collaboratively, they were informed that the grade they would get for the particular lab session (all the laboratory exercises determined 30% of their final course grade) would be formed equal parts determining the quality of their collaboration and the completeness and correctness of their joint solution. Dyads were arranged in space in a way that it was impossible for the students to use any other means of communication apart from these provided by Synergo, practicing thus the case of distant collaboration. The final dataset used in this study consisted of the collaborative sessions of 228 dyads.

4 Rating collaboration quality

Due to the limited evidence of the value of automatic metrics of interaction for indicating important aspects of collaboration, a statistical approach was followed that aims at comparing the information provided by these metrics with quantitative assessments of collaboration from another methodological standpoint. For that reason, a rating scheme approach was followed that involves human agents assigning ratings of collaboration quality in several of its core dimensions. Apart from leading to quantitative results, suitable for integrated elaboration with the metrics’ values, this approach takes into account deeper aspects of collaboration than calculations on event logs can convey, at least from a first point of view.

A rating scheme or a rating scale is “a measuring instrument that requires the rater or observer to assign the rated object to categories or continua that have numerals assigned to them” (Kerlinger, & Lee 2000, p. 736, cited in Meier 2005). Rating schemes are discriminated from coding schemes in that they are used to make a judgement on a larger piece of data each time, and are based on the knowledge and the critical skill of the human agent that applies them, in contrast to coding schemes that demand from the coder to neutralise the process by following strictly defined rubrics (Kerlinger, & Lee 2000).

The conceptual framework for the definition of core aspects of collaboration quality to be rated for each case is influenced by the work of Meier et al. (2007). This framework defines the main dimensions of collaboration quality that were operationalised using a concept-oriented rating scheme, stating precise definitions

of the concepts that determine the rating grades, and providing information of the means of correctly applying the process (Guilford, 1954). For that reason, a handbook including anchoring examples and guidelines for the correct conduction of the rating process is provided (Meier 2005). Therefore, the rating approach is normative, i.e. it compares assessments to an exemplary case of desired collaboration quality.

Due to some significant changes in the setting of collaborative activities, the mediating tools, the profiles of the students, and the design of the task, the framework and the rating tool were generalized and adapted so as to be suitable for the settings of activities of this study (Kahrmanis et al., 2009). The resultant rating scheme consists of 7 dimensions of collaboration quality: *collaboration flow*, *sustaining mutual understanding*, *knowledge exchange*, *argumentation*, *structuring the problem solving process*, *cooperative orientation*, and *individual task orientation* (Kahrmanis et al., 2009). The adapted version of the scheme is depicted in Table I, which contains each dimension of the rating scheme related to the general aspect of collaboration quality that it belongs to.

Table I: Dimensions of collaboration quality as defined by the adapted version of the rating scheme

General aspect of collaboration covered	Dimension of the adapted version of the rating scheme
Communication	Collaboration flow
	Sustaining mutual understanding
Joint information processing	Knowledge exchange
	Argumentation
Coordination	Structuring the Problem Solving Process
Interpersonal Relationship	Cooperative orientation
Motivation	Individual task orientation

Each collaborative session was then rated using a Likert-like scale: {-2,-1,0,1,2}. The rating process was based on video-like reproductions of the activities facilitated by the Synergo's playback tool, and an adapted rating handbook that guided raters' decisions (Kahrmanis et al, 2009). One rating was assigned for each session and dimension of the rating scheme. Two raters with prior experience with the task were responsible for the ratings, which were tested for inter-rater reliability using 33% of the dataset. Reliability scores were good: ICC ranged between .83 and .95, adjusted ICC between .84 and .95, Cronbach's alpha between .91 and .98, depending on the dimension rated (Kahrmanis et al, 2009). The resultant rated dataset was then ready for correlational analysis with the automatically calculated values of the metrics developed.

5 Correlation between metrics of interaction and collaboration quality

All statistical correlations between each metric and the ratings of collaboration quality for each dimension of the rating scheme were calculated. Some indicatory results are provided in Table I.

Table II: correlations between six metrics and dimensions of the rating scheme (including the average and absolute difference of the ratings in individual task orientation of the two participants and the average of the six first dimensions of the scheme). Upper value: Kendall's τ coefficient, and lower Spearman's ρ coefficient (distributions not normal)

	Coll. flow	Sust. mut. unders.	Knowl. exch.	Argumentation	Struct. Probl. Solv. Proc.	Cooper. Orient.	Ind. Task Orient. (avg)	Ind. task Orient. (abs. diff.)	Avg 6 dim.
C1: # of chat messages	.389**	.317**	.411**	.410**	.308**	.366**	.299**	-.211**	.409**
	.507**	.421**	.534**	.529**	.409**	.473**	.388**	-.270**	.564**
C4: altern. chat mes.	.407**	.349**	.406**	.427**	.309**	.389**	.327**	-.247**	.425**
	.527**	.460**	.535**	.552**	.414**	.503**	.420**	.314**	.583**
C6 mean res. time in chat	-.351**	-.304**	-.372**	-.379**	-.308**	-.330**	-.284**	.213**	-.380**
	-.464**	-.405**	-.489**	-.494**	-.413**	-.439**	-.374**	.274**	-.529**
MW1: # of MW actions	-.164**	-.102*	-.100*		-.116*				-.119**
	-.215**	-.135*	-.131*		-.153*				-.168*
MW3 symmetry in MW		.150**		.125*		.215**	.270**	-.340**	.112*
		.203**		.168*		.282**	.354**	-.433**	.166*
EV3 symm. in EV	.131**	.213**		.137**	.117*	.247**	.243**	-.320**	.152**
	.175**	.286**		.185**	.157**	.324**	.317**	-.410**	.220**

** p<0.01, * p<0.05

The main findings of this analysis are summed up in the following: chat-based metrics are highly correlated with all dimensions of collaboration quality. The highest correlations were found for the *collaboration flow* dimension and the dimensions that indicate information processing (*knowledge exchange* and *argumentation*). The most valuable chat-based metrics for indicating collaboration quality were the *number of chat messages*, the *alternation of chat messages* and the *mean response time in chat messages*. Almost all chat-based metrics correlated at statistically significant levels with most dimensions of the rating scheme. A notable exception was the *symmetry of chat messages* which did not correlate with any of the rating scheme's dimensions, since it took rather stable values throughout the dataset.

Concerning workspace-based metrics, one of the most notable findings relates to *symmetry in main actions* or *overall workspace actions*, which is a strong indicator of the difference in *individual task orientation* between participants, while it correlates at more moderate levels with *cooperative orientation*,

sustaining mutual understanding, and *argumentation* as well. On the other hand, metric MW1, the *volume of workspace-related actions*, is a negative indicator of collaboration quality on most of its dimensions and especially on the two communicational ones. The latter finding indicates that too much activity in the workspace is usually related to bad coordination and redundant actions in the workspace and hinders rather than aids communication and task coordination.

Overall events metrics usually convey similar information to the combination of metrics of the two distinct categories. In some cases, such as the total number of overall events metric (EV1), the two effects counteract and lead to non-statistically significant correlations with most dimensions. In other cases, however, this category provides metrics that provide additional information regarding their association with dimensions of collaboration quality. Such is the case with the symmetry in overall events (EV3), which correlates with dimensions of collaboration following approximately the same pattern with MW3 but with somewhat higher scores in most cases.

6 Conclusions and further research

This paper presented a study that designed and implemented a large set of metrics of interaction and examined the extent to which they can indicate core aspects of collaboration quality as the latter was defined and operationalised by Meier et al. (2007) and adapted by Kahrmanis et al. (2009).

Results of extended correlational statistical analysis in a large dataset of real-world collaborative activities revealed the extent of association of each metric with each distinct aspect of collaboration quality. Chat-based metrics were generally proved quite informative of desired collaboration practices or their absence, whereas workspace-based metrics provided insight into subtler issues of collaboration quality, reflecting negative facets of collaboration as well.

A respectable number of metrics was correlated at medium or high correlation scores with dimensions of collaboration quality, providing thus opportunities for the future development of models of automatic assessment of collaboration quality built on metrics reported above. Furthermore, findings obtained so far can inform the further refinement and development of automated metrics of interaction that belong to categories extensively correlated with dimensions of collaboration quality.

7 References

- Aditomo, A., & Reimann, P. (2007): 'Learning from virtual interaction: A review of research on online synchronous groups'. In C. Chinn, G. Erkens and S. Puntambekar (eds.): *Mice, minds and society, Proceedings of the 6th international conference on Computer Supported*

- Collaborative Learning, CSCL 2007*, International Society of the Learning Sciences, New Brunswick, NJ, 2007, pp. 56-65.
- Avouris, N., Dimitracopoulou, A., & Komis, V. (2003): 'On analysis of collaborative problem solving: An object oriented approach'. *Computers in Human Behavior*, vol. 19, no. 2, 147-167.
- Avouris N., Margaritis M., & Komis V. (2004): 'Modeling interaction during small-group synchronous problem-solving activities: "The Synergo approach"'. *2nd Int. Workshop on Designing Computational Models of Collaborative Learning Interaction, ITS2004*, Maceio, Brasil, September 2004.
- Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1995): 'The evolution of research on collaborative learning'. In P. Reimann and H. Spada (eds.): *Learning in humans and machines: Towards an interdisciplinary learning science*, Springer-Verlag, Berlin, pp. 189-211.
- Guilford, J. P. (1954). *Psychometric methods*. McGraw-Hill, New York.
- Kahrimanis G., Meier A., Chounta I., Spada H., Rummel N. and Avouris N. (2009): 'Assesing Collaboration Quality in Synchronous CSCL Problem-Solving Activities: Adaptation and Empirical Evaluation of a Rating Scheme', *LNCS, vol. 5794*, October 2009, Springer-Verlag, pp. 267-272.
- Kerlinger, F. N., & Lee, H. B. (2000): *Foundations of behavioral research*. Harcourt College Publishers, New York.
- Meier, A. (2005): 'Evaluating Collaboration: A Rating Scheme for Assessing the Quality of Net-Based, Interdisciplinary, Collaborative Problem Solving'. Unpublished diploma thesis, Institut für Psychologie, Albert-Ludwigs-Universität Freiburg im Breisgau, Freiburg, Germany.
- Meier, A., Spada, H. and Rummel, N. (2007): 'A rating scheme for assessing the quality of computer-supported collaboration processes'. *International Journal of Computer-Supported Collaborative Learning*, vol. 2, no 1, March 2007, pp. 63–86.
- Stahl, G. (2002): 'Rediscovering CSCL', In T. Koschmann, R. Hall and N. Miyake (eds.): *CSCL 2: Carrying forward the conversation*, Lawrence Erlbaum Associates, Hillsdale, NJ, 2002, pp. 169-181.